NL010449

Europäisches Patentamt | European Patent Office | Office européen des brevets

J1046 U.S. PTO
10/046472
01/14/02

# Bescheinigung | Certificate | Attestation

Die angehefteten Unterlagen stimmen mit der ursprünglich eingereichten Fassung der auf dem nächsten Blatt bezeichneten europäischen Patentanmeldung überein.

The attached documents are exact copies of the European patent application described on the following page, as originally filed.

Les documents fixés à cette attestation sont conformes à la version initialement déposée de la demande de brevet européen spécifiée à la page suivante.

**Patentanmeldung Nr. | Patent application No. | Demande de brevet n°**

01200142.6

Der Präsident des Europäischen Patentamts;
Im Auftrag

For the President of the European Patent Office

Le Président de l'Office européen des brevets
p.o.

I.L.C. HATTEN-HECKMAN

DEN HAAG, DEN
THE HAGUE, 25/10/01
LA HAYE, LE

EPA/EPO/OEB Form 1014 - 02.91

Europäisches Patentamt | European Patent Office | Office européen des brevets

# Blatt 2 der Bescheinigung
# Sheet 2 of the certificate
# Page 2 de l'attestation

Anmeldung Nr.:
Application no.: 01200142.6
Demande n°:

Anmeldetag:
Date of filing: 16/01/01
Date de dépôt:

Anmelder:
Applicant(s):
Demandeur(s):
Koninklijke Philips Electronics N.V.
5621 BA Eindhoven
NETHERLANDS

Bezeichnung der Erfindung:
Title of the invention:
Titre de l'invention:
Sinusoidal matching pursuit

In Anspruch genommene Prioriät(en) / Priority(ies) claimed / Priorité(s) revendiquée(s)

| Staat:<br>State:<br>Pays: | Tag:<br>Date:<br>Date: | Aktenzeichen:<br>File no.<br>Numéro de dépôt: |
|---|---|---|

Internationale Patentklassifikation:
International Patent classification:
Classification internationale des brevets:
/

Am Anmeldetag benannte Vertragstaaten:
Contracting states designated at date of filing: AT/BE/CH/CY/DE/DK/ES/FI/FR/GB/GR/IE/IT/LI/LU/MC/NL/PT/SE/TR
Etats contractants désignés lors du depôt:

Bemerkungen:
Remarks:
Remarques:

# Extended basis functions for sinusoidal matching pursuit algorithms

Bert den Brinker

Sinusoidal audio coding schemes assume the existence of so-called sinusoidal tracks. The track information is extracted from the original signal, stored or transmitted, and used for synthesis. The extraction of the track information is done by segmenting the signal and employing matching pursuit algorithms where the dictionary consists of truly sinusoidal patterns, i.e, locally (per segment) the tracks are modelled as constant-amplitude, constant-frequency sinusoidal patterns. It is proposed to extend these basis functions in matching pursuit algorithms to make a more powerfull and efficient dictionary.

## Things known so far

In a sinusoidal audio coding scheme, an audio signal is represented by sinusoidal components mainly. Typically, these components are extracted from the signal on a regular basis (constant update rate). For efficient coding, the frequencies, amplitudes and/or phases of sinusoidal components (which are evolving in time) in consecutive intervals are linked together such that differential coding of the frequencies and amplitudes can be applied. This is usually done by assuming that the tracks can be modelled as constant-amplitude, constant-frequency sinusoidal patterns in sufficiently small intervals [1, 2, 3, 4, 5, 6, 7].

This approach has recently been questioned. It was conjectured that not all relevant information could be extracted by the dictionary of sinusoidal patterns, and therefore it was proposed to use damped sinusoids [8, 9]. We note that, so far, it has not been conclusively established how relevant this extension is in practical situations and, furthermore, the newly introduced degrees of freedom (damping parameters) are contained in a nonlinear way in the error. This means that costly search procedures are required for the extraction of this information.

# The problem for which this invention brings the solution

In matching pursuit algorithms, one usually sees that constant-amplitude, constant-frequency sinusoidal patterns do not reduce the spectral peaks sufficiently: apparently the assumption of sinusoidal patterns per segment is violated. There are two possible consequences.

The first one is that the matching pursuit algorithm will model a single spectral peak by several distinct frequencies. In most cases, this is not necessary since the linking process (if applied properly) will induce a broadening of the spectral peaks due to amplitude or frequency glides. So in as far as a single peak is modelled by more than one frequency and the spurious peaks are not discarded by the psycho-acoustic model, they will more probably be a burden than a better description.

The second possible consequence is if a peak is modelled as a single sinusoid only, then the induced side-peaks by subtracting the sinusoidal pattern may constitute a problem. The cleaned signal is usually sent to another processing stage (e.g, residual or noise coder). Again, we get the problem that the linking introduces a broadening of the peak, and therefore the total effect is that parts of a spectral peak are modelled twice.

Since the model of constant-amplitude, constant-frequency localised patterns is apparently not sufficient, we propose to extend the basis functions such that these patterns cover all possible minor deviations from the assumption of stationarity. The first-order deviations can be modelled as linear amplitude and/or frequency glides. Furthermore, we want the extended basis to include patterns providing fair approximations to wave forms resulting from two sinusoidal tracks where the individual track information is lost due to the finite frequency resolution introduced by the segmentation. Lastly, we require that the extended basis does not bring about a large increase in computational complexity. The simplest way to do this is to consider degrees of freedom which appear linearly in the considered error signal.

## Embodiment

In order to prevent the estimation of spurious peak around an already estimated frequency, it is suggested to clean the spectrum in the neighbourhood of the estimated frequency. The simplest way of doing so is by estimating a polynomial expansion around the estimated harmonic, i.e. fitting not only

$$s_0(t) = A_0 e^{j\omega t} \tag{1}$$

but also

$$s_k(t) = A_k t^k e^{j\omega t} \tag{2}$$

with $k = 1, \cdots, K$. In this way, not only the Fourier transform of the approximation equals the original at a specific frequency but also the first $K$ derivatives at this frequency are equalised.

For real-valued signals this results in $2K$ patterns

$$s_k(t) \text{ and } s_k^*(t) \tag{3}$$

where * denotes conjugation and where we optimise over $\omega$ and $A_k$ ($k = 0, \cdots, K$). The parameters $A_k$ are all linearly involved in a quadratic optimisation (this in contrast to the proposal in [8, 9]) and the Grammian can be calculated analytically for each $\omega$.

Truncation of the expansion (i.e. the value of $K$) can be done independently for each spectral peak and can be made dependent on the convergence of the polynomial series.

The information of the higher order patterns can be used in the linking process. This information can be used to make more refined predictions of amplitude, frequency and phase of the considered sinusoidal in the next segment or, alternatively, can be used to steer the linking process.

Whether or not the extended data needs to be transmitted/stored has to be considered. This can again be done on basis the data itself. After having established the parameters and the track links, a psycho-acoustic model can determine the additional impact of the extra components on the sound quality. It is conjectured that in most cases this information need not be stored/transmitted but, rather, that is is more important to use the extended basis to prevent modelling of artifacts due to the improper modelling of already extracted sinusoidal components.

## Application areas

Sinusoidal audio coding

## References

[1] R. McAulay and T. Quartieri. Speech analysis/synthesis based on sinusoidal representation. *IEEE Trans. Acoust., Speech, Signal Process.*, 43:744–754, 1986.

[2] B. Edler. Technical description of the MPEG-4 audio-coding proposal from University of Hannover and Deutsche Bundespost Telekom. Technical Note MPEG95/0414, Int. Organisation for Standardisation ISO/IEC JTC1/SC29/WG11, 1995.

[3] K.N. Hamdy, M. Ali, and A.H. Tewfik. Low bit rate high quality audio coding with combined harmonic and wavelet representations. In *Proc. 1996 Int. Conf. Acoust. Speech Signal Process. (ICASSP96)*, pages 1045–1048, Atlanta GA, 7-10 May 1996. IEEE, Picataway, NJ.

[4] M. Ali. *Adaptive signal representation with application in audio coding.* PhD thesis, Univ. of Minnesota, 1996. Pages 1–165.

[5] M.M. Goodwin. *Adaptive signal models: theory, algorithms, and audio applications.* PhD thesis, Univ. of California, Berkeley, 1997. Pages 1–259.

[6] X. Serra. Musical sound modeling with sinusoids plus noise. In C. Roads, S. Pope, A. Picialli, and G. De Poli, editors, *Musical Signal Processing*. Swets & Zeitlinger, 1997.

[7] S.N. Levine. *Audio representation for data compression and compressed domain processing*. PhD thesis, Stanford Univ. (CA), 1999. Pages 1–136.

[8] J. Nieuwenhuijse, R. Heusdens, and E.F. Deprettere. Robust exponential modeling of audio signals. In *Proc. SPS 98*, pages 143–146, Leuven, Belgium, 26–27 March 1998.

[9] K. Vos. Audio signal decompositions. Master's thesis, Delft University of Technology, Delft, The Netherlands, 1998.

# Appendix

In order to avoid spending bits/analysis effort on signal components caused by slight variations with respect to sinusoidal basis functions (e.g., small amplitude variations, small frequency sweeps), it may be worthwhile to extract not only pure sinusoids, but to remove such spurious activity as well. In that case, we will not spend any effort on estimating these side effects as sinusoidal components, nor will they become part of the input to the noise coder.

## Regression with products of harmonics and polynomials

Consider the signal $s(t)$ with $s \in \mathbb{R}$ and $t \in \mathbb{R}$. We want to provide an approximation $\hat{s}$ of this signal $s$ on the interval $(-T/2, T/2)$. We consider the case of modelling according to

$$\hat{s}(t) = \Re\{[a + bt + ct^2]e^{j\Omega t}\} \tag{4}$$

where $\Omega \in \mathbb{R}^+$ and $a, b, c \in \mathbb{C}$.

**Amplitude variations**

It is obvious that second-order amplitude variations can be handled. Suppose the true signal $s$ is given by $s(t) = K(t)\cos\{\omega_0 t + \phi\}$ with $k \in \mathbb{R}$. Around $t = 0$ we have the Taylor series expansion

$$s(t) = (A + Bt + Ct^2)\cos\{\omega_0 t + \phi\} \tag{5}$$

with $A = K(0)$, $B = K'(0)$ and $C = K''(0)/2$. Consequently, the model approximation can be taken as $\Omega = \omega_0$, $a = Ae^{j\phi}$, $b = Be^{j\phi}$, $c = Ce^{j\phi}$. In order that the Taylor series expansion constitutes an accurate approximation, it is required that $K$ is approximately band limited with cut-off frequency $2\pi/T$. We note that all complex amplitudes $a, b, c$ are in phase.

**Frequency variations**

Consider (small) linear frequency sweeps:

$$s(t) = A\cos\{[\omega_0 + \Delta_\omega t]t + \phi\} \tag{6}$$

then we have as an approximation

$$
\begin{aligned}
s(t) &= \Re\left\{Ae^{j\phi}e^{j\omega_0 t}e^{j\Delta_\omega t^2}\right\} \\
&\approx \Re\left\{Ae^{j\phi}e^{j\omega_0 t}[1+j\Delta_\omega t^2+\cdots]\right\} \\
&= \Re\left\{[a+ct^2]e^{j\Omega t}\right\} = \hat{s}(t)
\end{aligned} \tag{7}
$$

with $\Omega = \omega_0$, $a = Ae^{j\phi}$, and $c = A\Delta_\omega e^{j\phi+\pi/2}$. For the approximation to hold it is required that

$$\Delta\omega(T/2)^2 \ll 1. \tag{8}$$

We note that $a$ and $c$ are out of phase.

**Non-resolvable frequencies**

Consider the signal

$$s(t) = A_1\cos\{\omega_1 t+\phi_1\} + A_2\cos\{\omega_2 t+\phi_2\} \tag{9}$$

with $|\omega_1-\omega_2| < 2\pi/T$ and $A_1, A_2 \in \mathbb{R}$. Without loss of generality, we consider $\omega_1 < \omega_2$. Furthermore, we define $\omega_0$ with $\omega_1 < \omega_0 < \omega_2$ and

$$
\begin{aligned}
\omega_1 &= \omega_0+\Delta_1, \\
\omega_2 &= \omega_0+\Delta_2, \\
B_1 &= A_1e^{j\phi_1}, \\
B_2 &= A_2e^{j\phi_1}.
\end{aligned}
$$

We note that $\Delta_1$ and $\Delta_2$ have opposite signs. We then have

$$
\begin{aligned}
s(t) &\approx \Re\left\{B_1e^{j\omega_0 t}(1+j\Delta_1 t-(\Delta_1)^2t^2/2+\cdots)+B_2e^{j\omega_0 t}(1+j\Delta_2 t-(\Delta_2)^2t^2/2+\cdots)\right\} \\
&= \Re\left\{[(B_1+B_2)+j(B_1\Delta_1+B_2\Delta_2)t-(B_1\Delta_1^2+B_2\Delta_2^2)/2t^2]e^{j\omega_0 t}\right\} \\
&= \Re\left\{[a+bt+ct^2]e^{j\Omega t}\right\} = \hat{s}(t)
\end{aligned} \tag{10}
$$

with $\Omega = \omega_0$, $a = B_1+B_2$, $b = j(B_1\Delta_1+B_2\Delta_2)$ and $c = -(B_1\Delta_1^2+B_2\Delta_2^2)/2$. Note that the choice of $\omega_0$ determines the approximation coefficients $b$ and $c$.

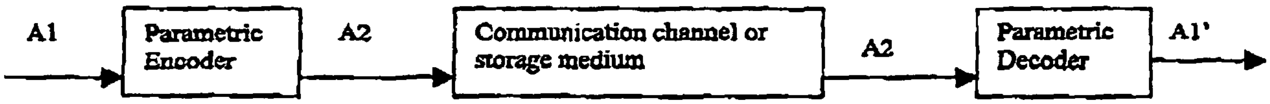


Fig. 1

Fig. 1 shows an embodiment of the invention. An audio and/ or speech signal A1 is furnished to a parametric encoder and coded into an encoded audio and/or speech signal A2. The encoded signal A2 is transmitted over a communication channel or stored on a storage medium. A parametric decoder obtains the encoded signal from the communication channel or storage medium and decodes this signal A2 into a decoded audio and/or speech signal A1' which is a representation of A1. The parametric encoder according to this embodiment of the invention extracts track information from A1 by employing a matching pursuit where the dictionary comprises extended basis functions as described above. Information on the relevant extended basis functions may be included in the bit-stream A2 and transmitted to the decoder. In the decoder, on the basis of the information present in the bit-stream A2, a reconstruction of the original audio signal is made: A1'. In this reconstruction in the decoder, the information in A2 on the extended basis functions may be used.

CLAIMS:

1. A parametric coding method of encoding an audio (and/ or speech) signal, which method comprises the step of extracting track information from the audio signal by employing a matching pursuit algorithm wherein the dictionary comprises extended basis functions.

2. A parametric encoder for encoding an audio (and/ or speech) signal, which device comprises means for extracting track information from the audio signal by employing a matching pursuit algorithm wherein the dictionary comprises extended basis functions.

3. A parametric decoding method of decoding an encoded audio (and/ or speech) signal, which method comprises the step of receiving the encoded audio signal which includes information on relevant extended basis functions, and using the information on the relevant extended basis functions in the reconstruction of an audio signal.

4. A parametric decoder for decoding an encoded audio and/ or speech signal, which decoder comprises means receiving the encoded audio signal which includes information on relevant extended basis functions, and means for using the information on the relevant extended basis functions in the reconstruction of an audio signal.

5. An encoded audio and/ or speech signal, which signal includes information on relevant extended basis functions.